

USING SOFT COMPUTING TOOLS FOR PIEZOMETRIC LEVEL PREDICTION

Joaquim Tinoco^{*}, Mathilde de Granrut^{†‡}, Daniel Dias[‡], Tiago Miranda^{*} and Alexandre-Gilles Simon[†]

^{*} ISISE - Institute for Sustainability and Innovation in Structural Engineering, School of Engineering, University of Minho
Campus de Azurém, 4800-058 Guimarães, Portugal
e-mail: jtinoco@civil.uminho.pt, webpage: <https://www.isise.net>

Keywords: Dam monitoring, Concrete Dams, Piezometric Levels, Data Mining techniques

Abstract. *The safety assessment of dams is a complex task that is made possible thanks to a constant monitoring of pertinent parameters. Once collected, the data is processed by statistical analysis models in order to describe the behaviour of the structure. The aim of those models is to detect early signs of abnormal behaviour so as to take corrective actions when required. Because of the uniqueness of each structure, the behavioural models need to adapt to each of these structures, thus flexibility is required. Simultaneously, generalisation capacities are sought, so a trade-off has to be found. This flexibility is even more important when the analysed phenomenon is characterised by non-linear features, as it is the case for the piezometric levels (PL) monitored at the rock-concrete interface of the arch dam that this study focuses on. In that case, the linear models that are classically used by engineers show insufficient performances. Consequently, interest naturally grows for the advanced learning algorithms known as machine learning techniques. In this work, the aim is to compare the predictive performances and generalization capacities of three different Data Mining algorithms that are likely to be used for monitoring purposes: Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Multiple Regression (MR). The achieved results show that SVM and ANN stand out as the most efficient algorithms, when it comes to analysing non-linear monitored phenomenon. Through a global sensitivity analysis, the influence of the models' attributes was measured, evidencing a high impact of Z (relative trough) in PL prediction.*

1 INTRODUCTION

Assessing the safety of dams is a priority for the owners of those large civil engineering structures. They are required to have a clear vision of the state of health of their dams, and to be able to potentially detect any abnormal evolution. In case such an event occurs, identifying the causes and taking the necessary steps to bring the structure back to a safe state is made possible thanks to a good knowledge and understanding of the behaviour of the structure.

[†] Électricité de France - Division Technique Générale (EDF DTG), Grenoble, France

[‡] 3SR Laboratory, Grenoble Alpes University, CNRS, Grenoble INP, Grenoble, France

Considering the specificity of each structure, being able to assess at any moment the state of a given dam is a complex task that relies on a systematic surveillance of the structure. This surveillance is based on the one hand on visual inspection, and on the other hand on monitoring. While visual inspections are rather qualitative techniques, the monitoring of dams is based on the continuous gathering of pertinent measurements that are processed by behavioural analysis models. The type of data that is collected is varied, including mechanical (displacements) and hydraulic quantities (piezometric levels, leakage flows), and constitute representative factors that traduce the global behaviour of the dam. In engineering practices, this behaviour is assumed to be simultaneously influenced by three external loads, namely the hydrostatic load, the thermal load, and the temporal load. Eventually, behavioural models are built, based on the measurements of both loads and their effects^{1,2,3,4}. The first aim of those models is to provide a prediction of the behaviour of the structure under normal operating conditions, which is compared to the actual measurements and make it possible to check the appropriateness between the expected and actual evolutions. Second, in a long-term perspective, sensitivity analyses are carried out to identify the contribution of each load to the considered effect.

Today, as most dams have been monitored since their first filling and with the generalisation of telemetry, a great amount of data is already available, which makes it possible to use statistical models. Among the community of dam owners, the classically used models belong to the category of the multi linear regression, and the reference model is the HST (Hydrostatic, Season, Time) model⁵. Initially designed to describe mechanical phenomena, this type of model assumes that the explanatory factors have independent and thus additive effects on the modelled quantity. Thus, its application to the analysis of hydraulic phenomenon is not always pertinent, for non-linearity comes into play and the additivity assumption is not confirmed. In order to deal with such non-linear phenomenon, more advanced models issued from the Data Mining (DM) techniques turn out to be particularly interesting, by providing valuable processing of the database.

The state of compression of the rock-mass foundation situated right under the contact between rock and concrete is in constant evolution, due to the variations of the abutment forces that the foundations support. The appearance of tensile stress is regularly observed at the heel of the dam which causes the permeability of the rock-mass to increase and the hydrostatic load is thus transferred to the foundation. The tensile stress can also induce an unsticking of the rock-concrete contact, and/or a cracking of the upstream face concrete⁶. This phenomenon is referred to as the opening of the rock-concrete interface. The aperture of the contact induces the rise of the piezometry in that zone, also known as the development of uplift pressure. Subsequently, it is possible to assess the state of aperture by interpreting the piezometric levels (PL) measured at the interface. Its temporal evolution is multi-scale. Indeed, the size of the aperture varies at the infra-annual scale, evolving with the mechanical stress that the dam is subject to, but its magnitude can also evolve at the scale of several years, with the opening and thus the full charge propagating further toward the toe of the dam, or on the contrary, declining subsequently to specific operational conditions.

Unlike most mechanical phenomenon, the aperture of the interface follows non-linear evolution rules. Indeed, because of the thermal sensitivity of concrete, the influence of a given filling, and thus a given hydrostatic load, will differ according to the thermal state of the structure: low temperatures would cause concrete to contract, inducing a global downstream movement of the arch, which exacerbates the tensile stress and finally increases the aperture. Because of those non-linear features, it is not possible to obtain a satisfying modelling of the PL at the interface by using mere additive models. This has led engineers to be interested in more advanced models in order to serve monitoring purposes. Accordingly, the following work aims at comparing the numerical performances of three different DM

techniques. Eventually, the interpretability of those complex algorithms is declined thanks to a global sensitivity analysis procedure.

2 CASE STUDY

The data that was proposed for that study comes from a French double curvature arch dam, which is 130-meter-high from the foundation to the crest, with a 425-meter-long crest. Its thickness varies from 25m to 6m, it is thus considered a thin arch. The ratio between the width of the valley (L) and the height of the dam (H) is higher than 3 ($L/H=3.3$), which indicates that the valley is relatively large. Those characteristics are elements that are known to favour the appearance of an aperture at the interface. Coupled to the phenomena of shrinkage and creep of concrete and the creep of the foundation, the arch shifted to the downstream direction right from the first filling of the reservoir, and the rock-concrete interface opened (5 to 7 mm). Subsequently, uplift pressures propagated under the heel of the dam, and further toward downstream. In order to follow the evolution of this phenomenon, the network of piezometers has been gradually complemented. It now comprises 52 piezometers, distributed under the dam and under the downstream plunge basin. Among those sensors, four are particularly interesting (named C1, C2, C3 and C4), since they are situated (Figure 1) in the central cantilever of the dam, at the rock-concrete interface (298 m). Therefore they are directly influenced by the aperture of this interface. The data provided for this study corresponds to the C4 sensor, situated close to the downstream end of the toe.

C4 is the piezometer that is situated at the smallest distance from the downstream end. Therefore, assuming that the aperture extends until the toe when the dam is submitted to extremely high stresses, C4 is situated in a zone where the interface is sometimes closed (for low fillings and/or high temperatures) and sometimes open (for high water levels and/or low temperatures). Thus, C4 is the most sensitive to the loads variations, and can be used to get an idea of the evolution of the aperture.

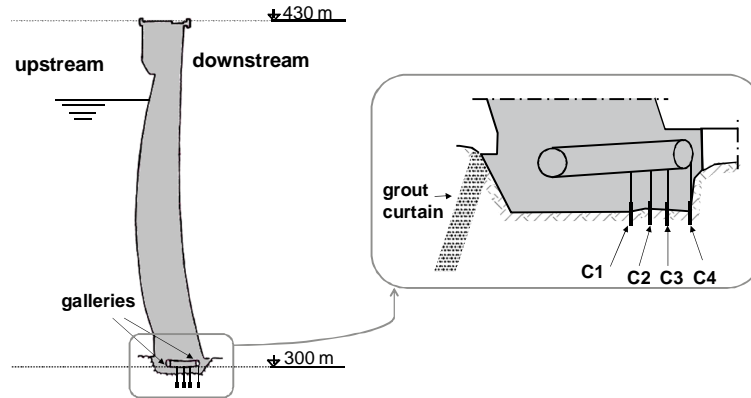


Figure 1: Piezometers location (central cantilever)

The time series that are provided stretch from September 2011 to June 2016 and comprise 623 observations for each measured quantity. The measured quantities are the following, with $i \in \{1, 623\}$ (all measurements are synchronised):

- the piezometric levels PL_i expressed in meters of a water column
- the water level in the reservoir h_i expressed in meters
- time t_i expressed in number of days elapsed since 1 January 2011 (the chosen origin)

The following variables are subsequently defined from those measurements:

- the season S_i which is an angle equal to 0° on the 1st of January and 360° on the 31st of December, defined by $S_i = 2\pi \cdot (t_i/365.25 - \text{floor}(t_i/365.25))$
- the relative trough Z_i which is a scaling of the water level h_i defined by $Z_i = (h_{\text{norm}} - h_i) / (h_{\text{norm}} - h_{\text{emp}})$, where h_{norm} is the normal operating water level and h_{emp} the water level when empty

In order to have a temporal distribution that would be as balanced as possible, a time sampling was processed so as to keep maximum one measurement per day. Indeed, in standard operating conditions, the sensors are automatically polled once a week during the hot season and/or for low water level, and once or twice a day during cold season and for high water levels. However, during some singular operational events, the data acquisition frequency increases in order to conduct the operation as safely as possible and follow its evolution closely. For the considered sensor, the interval between two measurements is often inferior to two days, and up to ten measurements per day are regularly observed (twice a year, the drainage system is locally closed and opened for efficiency reasons). Consequently, those very close measurements are correlated between each other, and the convergence of the models might be overly influenced by those identical observations. What's more, the fact that those dense observations correspond to singular operational conditions might result in a deterioration of the representativeness (keeping in mind that the models aim at describing the behaviour of the dam under normal operating conditions). Consequently, it was decided to keep only one measurement per day which resulted in 623 remaining observations (starting from an initial 1041-large dataset). No more advanced sampling was carried out, in order to keep a sufficient amount of observations.

3 MODELLING

3.1 The predictors

The modelling techniques that are compared for regression purposes are based on complex mathematical algorithms and do not take into account any physical law. Thus the choice of the input variables is a way to involve some physical understanding of the phenomenon at stake. In the present case, the algorithms are the tools that are used to model the link between the PL and the three main external loads. Each of these loads has an effect on the PL, and it is those effects that the models are expected to build adequately from the appropriate predictors, detecting non-linear interactions between the inputs. The first load is the hydrostatic load, and the corresponding predictor is the relative through Z . The second load is the thermal load, which includes the different annual thermal waves that induce cyclical temperature variation in the concrete of the structure. This global concrete temperature variation is the sum of different thermal variations (air temperature as the most influent but also water temperature, solar radiations, presence or absence of wind, temperature of the foundations etc, with potentially different phases), that can thus be modelled thanks to periodic functions of the season, with a one-year-long period. The corresponding predictors are thus $\cos(S)$ and $\sin(S)$. The last influence quantity is time, which is introduced into the regression through the variable t (days). What is ideally expected from a model is to offer the necessary flexibility to be able to provide good predictions without having to impose physical laws *a priori*. Indeed, when modelling a complex phenomenon, the user is likely to identify a new potentially influencing variable, which he might want to add as a predictor, but without necessarily having a precise idea of how it affects the modelled quantity. Thus, the model is expected to build the interactions between this predictor and the modelled quantity automatically. That is the reason why it was chosen in this study to keep the inputs as basic as

possible: Z , t , $\cos(S)$ and $\sin(S)$, in order to identify the algorithm that has the highest flexibility and adaptability.

3.2 DM algorithms

Considering the insufficient performance of the linear models that are classically used, this work intends to explore the capabilities of advanced statistical analysis, also known as DM techniques. For that, three different DM algorithms were applied to analyze piezometric data monitored on a French large arch dam: Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Multiple Regression (MR).

ANNs is a computational model based on the structure and functions of biological neural networks⁷. The information is processed using iterations among several neurons. ANNs are considered nonlinear statistical data modelling tools where the complex relationships between inputs and outputs are modelled or patterns are found. This technique is capable of modelling complex non-linear mappings and is robust to explore data with noise. In this study the multilayer perceptron that contains only feedforward connections, with one hidden layer containing H processing units called neurons, was adopted. Because the network's performance is sensitive to H (a trade-off between fitting accuracy and generalisation capability), a grid search of $\{0; 2; 4; 6; 8\}$ was adopted during the learning phase to find the best H value. Such grid search only considered training data, dividing it randomly into fitting (70%) and validation data (30%), where the validation error was used to select the best H . After selecting the best H value, the ANN is retrained with the whole training data. The neural function of the hidden nodes was set to the popular logistic function $1/(1 + e^{-x})$.

SVMs are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. They were initially proposed for classification tasks⁸. Then it became possible to apply SVM to regression tasks after the introduction of the ε -insensitive loss function⁹. The main purpose of the SVM is to transform input data into a high dimensional feature space using non-linear mapping. The SVM then finds the best linear separating hyperplane, related to a set of support vector points, in the feature space. This transformation depends on a kernel function. In this work the popular Gaussian kernel was adopted. In this context, its performance is affected by three parameters: γ , the parameter of the kernel; C , a penalty parameter; and ε , the width of a ε -insensitive zone¹⁰. The heuristics proposed by Cherkassky and Ma¹¹ were used to define the first two parameter values, $C=3$ and $\varepsilon = \hat{\sigma}/\sqrt{N}$, where $\hat{\sigma} = 1.5/N \cdot \sum_{i=1}^N (y_i - \hat{y}_i)^2$, y_i is the measured value, \hat{y}_i is the value predicted by a 3-nearest neighbor algorithm and N is the number of examples. A grid search (similar to the one used for ANN) of $2^{\{-1;-3;-7;-9\}}$ was adopted to optimize the kernel parameter γ .

In MR, several independent variables are linearly combined to predict the dependent (output) variable¹³. Due to its additive nature, this model is easy to interpret and is widely used in regression tasks. However, one of its main limitations relies on modelling problems of a non-linear nature. MR was essentially used in this study as a baseline comparison.

All experiments were conducted using the R statistical environment¹⁵ and supported through the rminer package¹⁶, which facilitates the implementation of several DM algorithms, as well as different validation approaches such as cross-validation.

3.3 Models assessment

For models evaluation and comparison three metrics commonly used in regression problems were used: mean squared error (MSE), root mean square error (RMSE) and the squared correlation coefficient (R^2). Typically, the lower the error, the better the predictive model, with zero corresponding to the highest model performance. However, while low

values of MSE and RMSE should be interpreted as indicating high model predictive capacity, R^2 should be as close as possible to one. When compared to MSE, RMSE penalizes more heavily a model that produces high errors in a few cases. Additionally, for a quick comparison of the different regression models, the regression error characteristic (REC) curve proposed by Bi and Bennett¹⁷ was built, which plots the error tolerance on the x-axis versus the percentage of points predicted within the tolerance on the x-axis.

The models generalization performance was accessed by 5 runs under a cross-validation (k-fold = 3) approach, where the data (P) are randomly sampled into k mutually exclusive subsets (P_1, P_2, \dots, P_k), with the same length¹³. Under this scheme, all of the data are used for training and testing. Yet, this method requires approximately k (the number of subsets) times more computation, because k models must be fitted. In addition, the three prediction metrics are always computed on test unseen data (as provided by the 3-fold validation procedure).

For models interpretability, a novel visualization approach based on sensitivity analysis (SA) proposed by Cortez and Embrechts¹⁸ was applied in this work. SA is a simple method that is applied after the training phase and measures the model responses when a given input is changed, allowing the quantification of the relative importance of each attribute as well as its average effect on the target variable. In particular, the Global Sensitivity Analysis (GSA) method¹⁸ was applied, which is able to detect interactions among input variables. This is achieved by performing a simultaneous variation of F inputs. Each input is varied through its range with L levels and the remaining inputs fixed to a given baseline value. In this work was adopted the average input variable value as a baseline and set $L=12$. With the sensitivity response of the GSA, two important visualization techniques were computed. First the input importance barplot was built, which shows the relative influence (R_a) of each input variable in the model. To measure this effect, the gradient metric (g_a) for all inputs was calculated. After that, the relative influence was computed.

$$R_a = g_a / \sum_{i=1}^I g_i \cdot 100 (\%) \quad \text{where,} \quad g_a = \sum_{j=2}^L |\hat{y}_{a,j} - \hat{y}_{a,j-1}| / (L - 1) \quad (1)$$

where a denotes the input variable under analysis, I is the number of input variables and $\hat{y}_{a,j}$ is the sensitivity response for $x_{a,j}$. Second, in order to analyze the average impact of a given input in the fitted model, the Variable Effect Characteristic (VEC) curve was used. For a given input variable, the VEC curve plots the attribute L level values (x-axis) versus the SA responses (y-axis). Between two consecutive $x_{a,j}$ values, the VEC plot performs a linear interpolation. To enhance the visualization analysis, several VEC curves can be plotted in the same graph. In such case, the x-axis is scaled (e.g. within [0,1]) for all x_a values.

4 RESULTS ANALYSIS AND DISCUSSION

Table 1 compares the performance of the three DM algorithms in PL prediction based on MSE, RMSE and R^2 metrics (mean value and respective 95% level confidence intervals according to a t-student distribution). Apart from MR all algorithms present a very good response in PL prediction, with a R^2 very close to one. The highest performance in PL prediction was achieved by the ANN model, with an $R^2 = 0.9912$, very closely followed by the SVM ($R^2 = 0.9859$).

Figure 2a compares REC curves of all models confirming the poor performances of MR and highlighting the high accuracy of the ANN and the SVM. In fact, both models are able to predict around 96% of all records with an absolute deviation lower than 5 m. Even for a tighter tolerance (e.g. 2.5 m), ANN presents an accuracy higher than 85%. Figure 2b depicts

the relation between observed and predicted PL values (scatterplot) according to ANN model, showing once again a very interesting fit (all points are very close to the diagonal line).

Model	MSE	RMSE	R ²
ANN	3.91±0.53	1.97±0.13	0.9912±0.0012
SVM	6.57±0.92	2.56±0.18	0.9859±0.0019
MR	125.38±6.99	11.20±0.31	0.7174±0.0143
HST	56.42	7.51	0.8726

Table 1: Models performance comparison based on MSE, RMSE and R² metrics.

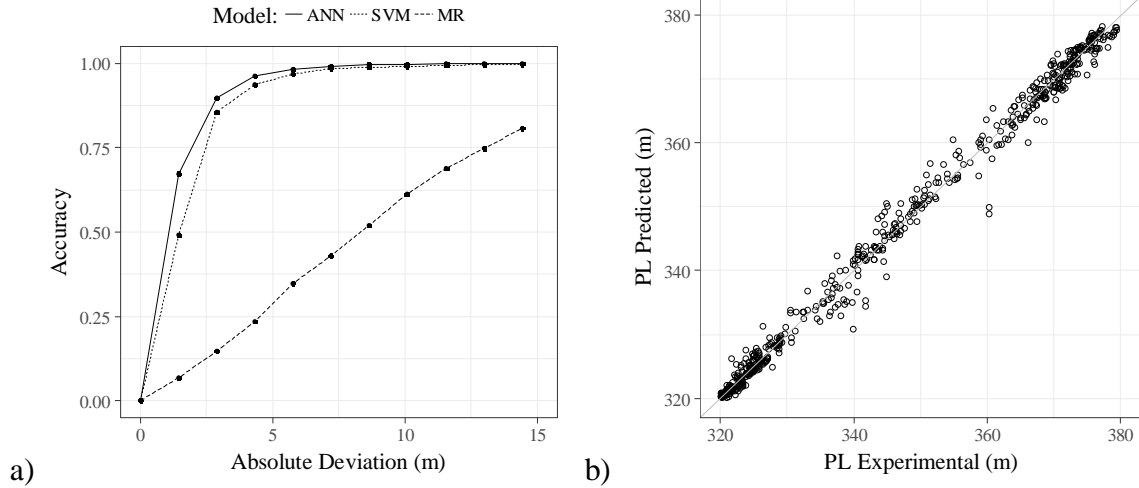


Figure 2: Models performance: a) REC curves of all models; b) Scatterplot of ANN model

On Figure 3a is plotted the relative importance of each input according to the three DM models. Note that since those inputs were determined in order to represent the three particular external loads (hydrostatic, thermal and temporal) that affect the behaviour of the dam, which is why $\cos S$ and $\sin S$ have to be interpreted jointly. From its analysis, there is no doubt that Z is the most relevant variable in PL prediction, which is confirmed by the three algorithms. Taking ANN model as reference, Z has a relative influence around 50%, followed by $\sin S$ and $\cos S$ with around 36%. Time is the least influencing variable. This plot (Figure 3a) permits to classify the inputs relatively compared to each other, however it has to be considered with caution because it does not in any case assess to what absolute extent they explain the PL. The purpose is not to select or leave inputs out of the analysis. The main interest of this analysis is that it shows that the studied PL are very sensitive to the water level variations. Since those PL are an image of the aperture of the rock-concrete interface, it means that the size of the aperture is very much correlated with the annual variations of the water level. This is in accordance with engineering knowledge and gives even more credit to the model. From an engineer point of view, this also shows that a way to control the aperture and limit its expansion is to adapt the water level.

In terms of physical behaviour, this relative importance analysis shows that the hydrostatic load plays a dominating role in determining the state of stress of the dam. The thermal load comes next, and although time is the least significant influence quantity, it does not mean that it should be removed from a diagnostic analysis.

Although the effects of the thermal and the hydrostatic loads are quantitatively the most significant effects, they correspond to elastic evolutions, which imply that it does not induce permanent deformation of the structure. For instance, if a high water level implies a significant movement downstream, simply lowering the water level will permit to have the

dam come back to a safe position. Consequently, the effects of the hydrostatic and the seasonal loads in usual ranges are not issue of too great concern. This is not the case for the irreversible evolutions that appear as time passes. In the context of dam safety, those temporal evolutions have to be identified and explained as early as possible, in order to conduct maintenance operations if needed.

Since ANN model achieved one of the highest performances, the analysis was pushed further for that algorithm. In order to get a better understanding of the PL and of the global behaviour of the structure, the effect of each load on the PL prediction was measured based on a SA¹⁸. In order to draw some physical conclusions on each model, the SA was performed reasoning in terms of load and not in terms of input. This means that to study the effect of the hydrostatic load, respectively the temporal load, two respective 1-D SA were performed, having only Z , respectively t , vary through its range. In order to study the impact of the thermal load, a 2-D SA was used, having both $\cos S$ and $\sin S$ vary simultaneously.

Accordingly, Figure 3b overlaps the VEC curves of the temporal load (t -variable), the hydrostatic load (Z -variable) and S the thermal load (S -variables). xx axis is scaled to accommodate all loads. Focusing first on the influence of the Z -variable, what is noticeable on the ANN predictions is that its effect is nearly non-existent when it varies between 0.5 and 1. However, when decreasing Z below 0.5, PL start to rise, and Z stands out as the most influencing quantity. Going back to the definition of the relative trough Z , one notices that Z is maximal when the water level h is minimal and vice versa. Thus, Figure 3b shows that when the water level in the reservoir increases, the PL at the interface increase as well, which is perfectly consistent with the state of the art. What is particularly interesting here is that the rise of the PL occurs only after a threshold is reached (0.5), which can be linked to the state of aperture of the rock-concrete interface. The interpretation of this non linearity is that this threshold water level corresponds to the moment at which the interface starts to open, leading to the hydrostatic load being transmitted to the foundations, and subsequently the rise of the PL. Once the aperture is open, the higher the water level the higher the PL. This non-linear feature that characterizes the evolution of the aperture of the rock-concrete interface is what made indispensable the use of models that are more advanced than additive models, such as HST.

Second, the curve corresponding to the influence of the thermal load traduces the sensitivity of the PL to the variations of the temperature of the body of the dam over the year, which follows a seasonal variation. For the ANN model, the corresponding curve has a quasi-sinusoidal shape, with the maximum being reached approximately at the first quarter of the year, which is April, and the minimum being reached in summer. The minimum is not perfectly clear though, because one point seems to distort the curve and draw a second maximum, though a smooth sinusoidal shape is expected. Observing the maximum predicted PL during the cold season is consistent with the engineering knowledge of how the thermal load influences the structure. Indeed, the thermal sensitivity of concrete induces a contraction of concrete which leads to a global downstream movement of the dam. This movement is conducive to the expansion of the rock-concrete aperture, which eventually results in the rise of the PL. However, because of the thermal inertia of concrete, there is a delay between the air temperature minimum, reached in average between January and February in the region where the dam is situated, and the maximum temperature of the body of the dam (including the foundations). Conversely, the minimum predicted PLs are observed during the hot season, which coincides with the moment of the year when the thermal state lets concrete expand and causes the upstream movement of the arch, minimizing the strain on the rock-concrete interface. Consequently, the aperture of the contact tends to close and eventually the PL decrease. Thus this sensitivity analysis confirms the validity of the ANN which describes accurately the response of the dam to the thermal load.

Finally, the curve corresponding to the time variable is also of great interest, because it shows that the PL decrease as t increases. In terms of structural behaviour, it means that PL decrease when time passes, and since the analyzed PL are directly linked to the aperture of the interface, this curve shows that the aperture is gradually closing on the period of analysis. From an engineer point of view, the closing of the interface is synonymous with an improvement of the behaviour of the whole structure, and thus an enhancement of its safety.

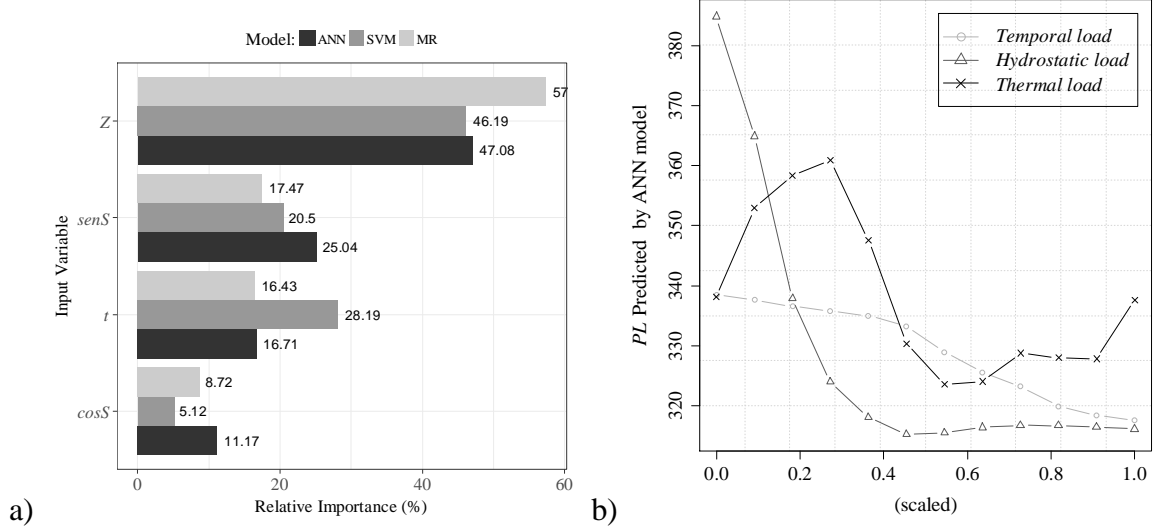


Figure 3: Models interpretation: a) Relative importance of each input variable; b) VEC curves of the temporal, hydrostatic and thermal loads according to the ANN model in PL prediction, based on a SA

5 FINAL REMARKS

This work proposed to challenge three Data Mining (DM) techniques in order to determine which of them was the most suitable to serve dam monitoring purposes. More particularly, the comparison was based on the analysis of piezometric measurements that were recorded at the rock-concrete interface of an arch dam, in order to get a better understanding of the phenomenon of the aperture of the interface. The three DM algorithms were fed with the same four basic inputs: time, the sine and cosine of the season, and the scale water level. From their comparison, ANN stood out as the most performing algorithm in terms of prediction, closely followed by SVM.

In order to draw conclusions on the behaviour of the structure, a sensitivity analysis (SA) was performed for ANN model, based on relative importance plots and Variable Effect Characteristic (VEC) curves. This SA was carried out in order to show the influence of the three external physical loads that affect the behaviour of a dam, namely the hydrostatic load, the thermal load and the influence of time. ANN was able to detect the coupled influence of the loads on the piezometric levels, and the observed effects were in accordance with the engineering knowledge. This work thus shows that ANN can describe more adequately the effect of the influencing loads. This leads to a better understanding of how the aperture of the rock-concrete interface evolves, which is a significant issue for dam safety.

Noticeably, DM techniques are flexible enough to build thresholds and combine the inputs so as to compute correct effects. Thus, suppose that one variable is approached for having an influence on a given phenomenon, DM algorithms can be used to add “blindly” this new input. The work of the engineer will then be to interpret the outputs of those algorithms to define how this input impacts the studied phenomenon. DM techniques can indubitably provide great improvement to the dam monitoring profession, but interpreting them falls within the competence of experienced engineers.

6 ACKNOWLEDGEMENTS

The authors thank the ANRT CIFRE for its grant (number 0902/2016) that partly supported this work.

REFERENCES

- 1 Carrère, Colson, Goguel, and Noret. (2000). “Modelling: a means of assisting interpretation of readings.” XXth International Congress on Large Dams, vol. III, Beijing, 1005–1037.
- 2 Léger, P., and Leclerc, M. (2007). “Hydrostatic, Temperature, Time-Displacement Model for Concrete Dams.” *Journal of Engineering Mechanics*, 133(3), 267–277.
- 3 Mendes de Vasconcelos Braga Farinha, M. L. (2010). “Hydromechanical behaviour of concrete dam foundations. In Situ tests and numerical modelling.”
- 4 Penot, I., Daumas, B., and Fabre, J.-P. (2005). “Monitoring behaviour.” *Water power and dam construction*, 57(12), 24:27.
- 5 Willm, and Beaujoint. (1967). “Les méthodes de surveillance des barrages au service de la production hydraulique d’Electricité de France, problèmes anciens et solutions nouvelles.” IXth International Congress on Large Dams, Istanbul, 529–550.
- 6 le Delliou, P. (2003). *Les barrages: conception et maintenance*. (P. U. Lyon, ed.).
- 7 Kenig, S., Ben-David, A., Omer, M., and Sadeh, A. (2001). “Control of properties in injection molding by neural networks.” *Engineering Applications of Artificial Intelligence*, 14(6), 819–823.
- 8 Cortes, C. and Vapnik, V. (1995). “Support vector networks.” *Machine Learning*, 20(3), 273–297.
- 9 Smola, A. and Schölkopf, B. (2004). “A tutorial on support vector regression.” *Statistics and 404 Computing*, 14(3), 199–222.
- 10 Safarzadegan Gilan, S., Bahrami Jovein, H., and Ramezaniapour, A. (2012). “Hybrid support vector regression–particle swarm optimization for prediction of compressive strength and rcpt of concretes containing metakaolin.” *Construction and Building Materials*, 34, 321–329.
- 11 Cherkassky, V. and Ma, Y. (2004). “Practical selection of svm parameters and noise estimation for svm regression.” *Neural Networks*, 17(1), 113–126.
- 12 Berry M, Linoff G. *Mastering data mining: the art and science of customer relationships management*. New York: John Wiley & Sons; 2000.
- 13 Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*. New York: Springer-Verlag; 2009.
- 14 Genuer, R., Poggi, J. M., Tuleau-Malot, C., & Villa-Vialaneix, N. (2017). *Random forests for big data*. Big Data Research.
- 15 Team, R. (2009). “R: A language and environment for statistical computing. R Foundation for Statistical Computing, Viena, Austria. Web site: <http://www.r-project.org/>.
- 16 Cortez, P., *Data Mining with Neural Networks and Support Vector Machines Using the R/rminer Tool*, in *proceedings of Advances in Data Mining - Applications and Theoretical Aspects 10th Industrial Conference on Data Mining (ICDM 2010)*, Lecture Notes in Artificial Intelligence 6171, 2010, pp. 572-583.
- 17 Bi, J., & Bennett, K. (2003). Regression error characteristic curves. In *Proceedings of the twentieth international conference on machine learning*. 43–50. AAAI Press, Washington.
- 18 Cortez, P., & Embrechts, M. (2013). Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences*, 225, 1–17.